Methods for the Production of a Channel-forming Protein

The invention relates to a method for the production of a channel-forming protein, a channel-forming protein, a gene encoding such a protein and mutated *mspA*, *mspC* or *mspD* genes, a plasmid vector and an overexpression system.

The invention relates in general to the technical field of the production of nanostructures. To date, the best characterised nanostructures are the carbon nanochannels (Yakobson, B. I. und Smalley, R. E. Fullerene nanotubes: $C_{1,000,000}$ and beyond. *Am Sci* **85**, 324, 1997). It was shown that the electronic properties of carbon nanochannels could be controlled through their structural details. Carbon nanochannels are synthesised using different variants of CVD (chemical vapour deposition) methods (Fan, S., Chapline, M. G., Franklin, N. R., Tombler, T. W., Cassell, A. M. und Dai, H. Self-oriented regular arrays of carbon nanotubes and their field emission properties. *Science* **283**, 512-4,1999), which therefore is very sumptuous.

From Johnson, S. A., Ollivier, P. J. and Mallouk, T. E. „Ordered mesoporous polymers of tuneable pore size from colloidal silica templates." *Science* **283**, 963-965 (1999) a technique for creating organic nanochannels on the basis of a template is reported. With this process, nanochannels with a diameter from 5 to 35 nm can be produced.

Mycobacteria belong to a subgroup of Gram-positive bacteria, which contain mycolic acids and include the genera *Corynebacterium, Nocardia, Rhodococcus, Gordona, Tsukamurella, Dietzia*.

Trias, J. and Benz, R. „Permeability of the cell wall of *Mycobacterium smegmatis*." *Mol Microbiol* **14**, 283-290 (1994) describe channel-forming proteins, called porins, in the mycolic

acid layer of mycobacteria. Biochemical or molecular genetic data of these porins have not been published yet.

From Lichtinger, T., Burkovski, A., Niederweis, M., Kramer, R. and Benz, R. „Biochemical and biophysical characterization of the cell wall porin of *Corynebacterium glutamicum*: the channel is formed by a low molecular mass polypeptide." *Biochemistry* **37,** 15024-32 (1998) the technique is known to prepare porins from corynebacteria. This technique is relatively inefficient, though.

Mukhopadhyay, S., Basu, D. and Chakrabarti, P. „Characterization of a porin from *Mycobacterium smegmatis*." *J Bacteriol* **179,** 6205-6207 (1997) describe the extraction of porins from *M. smegmatis* with a buffer containing 1 % Zwittergent by incubation at room temperature for one hour. The yields were poor and the porins were contaminated with many other proteins.

From Harth, G. et al., „High-level heterologous expression and secretion in rapidly growing nonpathogenic mycobacterium of four major *Mycobacterium tuberculosis* extracellular proteins considered to be leading vaccine candidates and drug targets." *Infection and Immunity* **65,** 2321-2328 (1997) it is known that a strong expression of *Mycobacteria*-specific proteins in *E. coli* seems to be not possible.

Senaratne, R.H. et al., „Expression of a Gene for a Porin-Like Protein of the OmpA Family from Mycobacterium tu*berculosis* H37Rv." *J Bacteriol* **180**, 3541-3547 (1998) describe the expression of a gene for a porin-like protein from *Mycobacterium tuberculosis* H37Rv in *E. coli*. The expression of the gene causes a discontinuance in the growth of the bacteria, evidently because the expressed protein is toxic for *E. coli*. A negligible amount of protein from *E. coli* could be isolated shortly before the dying of the cells.

It is an object of the invention, to provide an improved method for the production of a channel-forming protein.

This object is solved by the features of Claims 1, 26, 27, 28, 30, 32, 36, 37, 38, 40 and 41. Further suitable embodiments derive from the features of Claims 2 through 25, 29, 31, 33, 34, 35, 39, and, optionally, 37 and 38.

According to the invention, a method is provided for producing a channel-forming protein, found in Gram-positive bacteria, in which the channel-forming protein is produced through

a) heterologous overexpression or

b) purification from mycobacteria, wherein the extraction temperature is higher than 50°C.

When referring to channel-forming proteins, a protein is meant, which can form a water-filled channel or a water-filled channel-like structure. Such proteins occur naturally, especially in the cell walls of bacteria. They can form channel-like structures with diameter up to 3 nm or even larger. The length can be up to 10 nm or more. The channel-like structures or channels can be made up of many substructures, specifically 4 or 8 substructures.

This method according to the invention is much more efficient in comparison to the prior procedures, offers the possibility of a far-reaching automation of the chromatographic purification, and allows for a drastically increased yield.

The gram-positive bacterium can be a bacterium, which contains at least one mycolic acid. In the described According to one

embodiment, the bacterium is a mycobacterium, preferably *Mycobacterium smegmatis*.

The channel-forming protein can be a porin. Preferred is a porin, which is chemically stable in organic solution and/or thermally stable up to a temperature of 80°C, more preferably 100°C. Stable shall mean that the channel-like structure of the protein remains intact and essentially no denaturation of the protein occurs.

Preferred are the porins MspA, MspC, MspD, a fragment of these porins, a protein homologous to these porins or their fragments, or a protein of a sequence derived from these porins. MspA corresponds to the sequence of the amino acids 28 - 211 of sequence 3 (see below), MspC corresponds to sequence 7 and MspD corresponds to sequence 9. The homologous protein of said porins or their fragments exhibit a similar structure to that of said porin or their fragments. At least 20% of the amino acids are identical or homologous to the amino acids of these porins or fragments. An amino acid in a protein is homologous with another amino acid if it can be substituted with the other amino acid, without influencing the function or structure of the protein. A protein that has been deduced from a sequence of a porin can be missing a single or several amino acids when compared to the sequence, or contain other amino acids or amino acid analogues.

Said proteins are particularly suitable for the production of nanostructures because of their surprisingly high chemical and thermal stabilities.

A good yield will be obtained, if the heterologous overexpression is performed in *E. coli* or in mycobacteria. For overex-

pression, a gene encoding a channel-forming protein, preferably a porin, should be used. For overexpression, it is preferable to use an *mspA* gene according to sequence 1 (see below), an *mspC* gene according to sequence 6 or an *mspD* gene according to sequence 8. A mutant gene derived from the sequences 1, 6, or 8 can be used for overexpression, whereby the mutation is so that the chemical and thermal stability, as well as the channel-like structure, essentially correspond to that of MspA, MspC or MspD. The mutation can also be such that the codon usage of the *mspA*, *mspC* or *mspD* gene is adapted to that of highly expressed genes in *E. coli*. These codons are known from Nakamura, T. et al., „Two types of linkage between codon usage and gene-expression levels." *FEBS Lett.* **289**, 123-125 (1991).

A mutant *mspA*, *mspC* or *mspD* gene can also be used for overexpression, if the mutation essentially reduces the G+C content to less than 66%. The adaptation of the codon usage dramatically improves the overexpression of MspA, MspC and MspD in *E. coli*.

The yield of the channel-forming protein MspA can be further increased by a factor 10 to 20 through overexpression in *E. coli* compared to the method for preparation of the native protein described above.

It is useful to use the *synmspA* gene according to sequence 4 for overexpression. An overexpression vector for *E. coli*, in which the *synmspA* gene according to sequence 4 is inserted, can be used for this purpose. Suitable vectors are described by Hannig, G. and Makrides, S.C. in „Trends in Biotechnology", 1998, Vol. 16, pp54. The disclosure of this document is incorporated herein.

It has also been advantageously found to harvest the channel-forming proteins from the cell wall of the gram-positive bac-

teria by using non-ionic or zwitterionic detergents. The detergents can be selected from the following group: isotridecyl poly(ethylene glycolether)$_n$, alkyl glucosides, in particular octyl glucoside, alkyl maltosides, in particular dodecyl maltoside, alkyl thioglucosides, in particular octyl thioglucoside, octyl-polyethylenoxides and lauryl dimethylaminoxide. A twofold or higher critical micellar concentration (CMC) in a phosphate buffer (100 mM Na_2HPO_4/NaH_2PO_4, pH 6.5, 150 mM NaCl) is preferably used. The zwitterionic and non-ionic detergents very effectively dissolve the channel-forming protein MspA from the cell wall of *M. smegmatis*, resulting in a good yield.

It has further been shown as useful that the extraction temperature is between 80 and 110 °C, preferably between 90 and 100 °C and/or the extraction time is 5 - 120 min, preferably 25 - 35 min. Particularly preferred is the use of a buffer with an ionic strength of more than 50 mM NaCl or Na-phosphate.

In particular, performing the extraction at 100 °C and the use of a buffer with a high ionic strength or zwitterionic and non-ionic detergents will improve the method for extraction of porins from *Mycobacterium smegmatis*. In comparison with the prior procedures for purification of such proteins using organic solvents or their extraction at room temperature, it offers the following advantages:

aa) no use of organic solvents required
bb) minimal contamination with other proteins
cc) efficient extraction

It is also possible to purify MspA by dissolving it in dimethyl-sulfoxide at a temperature in the range from 50 - 110 °C; afterwards the solution is allowed to cool to room temperature, permitting filtration of the MspA precipitate.

Preferably, the channel-forming protein is precipitated, in particular using acetone, for purification. This procedure can result in the further concentration of MspA with respect to other non-precipitating proteins. It is also advantageous to purify the protein using an ion-exchange chromatography method, especially an anion-exchange chromatography method. Further purification can be achieved employing size-exclusion chromatography.

The renaturation of the channel-forming protein generated by means of heterologous overexpression, can be achieved by increasing the local concentration of the protein. The increase can be achieved using electrophoretic concentration, especially by means of a DC current, by precipitation or adsorption at a suitable surface (e.g. a membrane). Useful is a DC current of 50 V for 30 min.

Another aspect of the invention is a channel-forming protein from a gram-positive bacterium, produced according to the method of the invention.

The gram-positive bacterium can be a bacterium containing mycolic acids, whereby it is advantageous to use a mycobacterium, preferably *Mycobacterium smegmatis*.

It is especially advantageous that the channel-forming protein is a porin, which is chemically stable against organic solvents. The porin is essentially thermally stable up to a temperature of 80 °C, preferably up to 100 °C. This thermal stability is displayed by either MspA, MspC, MspD, a fragment of these porins, or a homologous porin to one of these fragments, or a porin which is derived from a sequence of the porins (MspA, MspC, MspD) or their fragments. The chemical and thermal stability, as well as the channel-forming structure of

such derived proteins, can in general correspond to the stabilities of the MspA, MspC, MspD proteins. It is further possible that additional proteins, which are not mentioned here, possess very similar properties and are therefore encompassed by the scope of the present invention.

The channel-forming proteins according to the present invention have the following advantages:

aaa) If the channel-forming proteins are found in the cell wall of *M. smegmatis*, they can be dissolved in organic solvents (e.g. $CHCl_3$/MeOH) without denaturation. The channel-forming property remains also in organic solvents.

bbb) They can be precipitated using acetone without denaturation.

ccc) They even survive being boiled in detergents (e.g. 10 min in 3% SDS) without denaturation.

This extreme stability of the inventive proteins against chemical and thermal denaturation makes it possible to use them in order to produce technically applicable nanostructures.

According to invention, a gene is furthermore claimed, which encodes a channel-forming protein according to the invention. This can be the *mspA* gene according to sequence 1, the *mspC* gene according to sequence 6 or the *mspD* gene according to sequence 8.

As an additional matter, a mutated *mspA* gene, *mspC* gene or *mspD* gene is provided, in which the codon usage of the aforesaid gene is practically the same as the codon usage of highly expressed *E. coli* genes. The mutation can be such that the G+C

content is reduced to less than 66%. The mutated gene can also be derived from one the sequences 1, 6, or 8 in such a way that the chemical and thermal stability, as well as the channel-like structure of the expressed protein, is essentially the same as that of MspA, MspC or MspD. Additional mutations that are not mentioned here are conceivable for the skilled artisan. Genes that lead to the formation of channel-like proteins according to the invention are herewith included in the scope of protection as claimed, e.g. a mutated *mspA* gene, in which the mutated gene is the *synmspA* gene according to sequence 4 (see below).

An additional object of the present invention is the plasmid vector pMN501 and an overexpression system, wherein *E. coli* contains said plasmid vector.

In the following examples of the invention are explained with reference to the figures. These show:

Fig. 1a-c the temperature dependent extraction of MspA from *M. smegmatis* as shown by gel electrophoresis,

Fig. 2 the purification of MspA from *M. smegmatis* as shown by gel electrophoresis,

Fig. 3 the purification of MspA from *E. coli* as shown by gel electrophoresis,

Fig. 4 the construction of plasmid vector pMN501,

Fig. 5 a scheme depicting an apparatus for renaturing monomeric MspA,

Fig. 6 renatured MspA as shown by gel electrophoresis and

Fig. 7a-c modifications of the channel-forming protein MspA as shown by electron microscopy.

The proteins were incubated at room temperature for 30 minutes in a sample buffer containing 40 mM tris(hydroxymethyl)aminomethane, pH 7.0, 3% sodium dodecyl sulfate, 8% glycerol, 0.1% Serva Blue G) in all gel electrophoretic experiments and then separated according to their sizes by gel electrophoresis.

The figures 1a-c show proteins extracted from *M. smegmatis* at different temperatures.

Fig. 1a. 10% denaturing polyacrylamide gel stained with Coomassie Blue. Lane M: molecular weight marker (200, 116.3, 97.4, 66.3, 55.4, 36.5, 31, 21.5 und 14.4 kDa). Lanes 1 through 8: 12 µl of each extract obtained at 30, 40, 50, 60, 70, 80, 90 and 100°C.

Fig. 1b shows an immunoblot analysis of an 8% denaturing polyacrylamide gel blotted onto a PVDF membrane. Proteins were visualized using an MspA antiserum and a chemoluminescence reaction (ECL detection system, AmershamPharmacia, Vienna, Austria). Lane M: molecular mass marker (97.4, 68, 46, 31, 20.1, 14.4 kDa kDa); lanes 1 through 3: 2 µL of extracts obtained at 30, 40 or 50 °C; lanes 4 through 8: 1 µL of extracts obtained at 60, 70, 80, 90 or 100 °C; (9) 1 ng MspA.

Fig. 1c shows a denaturing 8% polyacrylamide gel stained with silver. Lane M: molecular mass marker (200, 116.3, 97.4, 66.3, 55.4, 36.5, 31, 21.5, 14.4, 6 kDa kDa); lanes 1 and 2: 15 µL of extracts obtained at 30 and 40°C, respectively; lane 3: 10 µL of an extract obtained at 50°C, lanes 4 through 8: 4 µL of extracts obtained at 60, 70, 80, 90 or 100 °C; (9) 270 ng purified MspA.

Fig. 2 shows a denaturing 10% polyacrylamide gel stained with Coomassie Blue.

Lane M: molecular mass marker (200, 116.3, 97.4, 66.3, 55.4, 36.5, 31, 21.5, 14.4, 6 kDa kDa); Lane 1: 40 µg protein of an extract from *M. smegmatis* obtained using POP05 buffer (100 mM Na_2HPO_4/NaH_2PO_4, pH 6.5, 0.1 mM EDTA, 150 mM NaCl, 0.5 % octyl-polyethylenoxide (OPOE)). Lane 2: 40 µg protein of an extract after precipitation with acetone. Lane 3: 4 µg protein after anion-exchange chromatography and pooling of the fractions, which contained MspA. Lane 4: 4 µg protein of the pooled MspA fractions after precipitation with acetone. Lane 5: 4 µg protein after size-exclusion chromatography and pooling of the fractions, which contained MspA. The sequences of the *mspA* gene, of the *mspA* gene + promoter and the MspA protein with the putative signal sequence are shown as sequences 1 to 3 in the sequence protocol.

Fig. 3 shows the purification of the channel-forming protein MspA from *E. coli*. The proteins were separated according to their sizes in a 10% denaturing polyacrylamide gel. The gel was stained with Coomassie Blue. Lane 1: Lysate from *E. coli* BL21(DE3)/pMN501 before induction with IPTG. Lane 2: Lysate from *E. coli* BL21(DE3)/pMN501 after induction with IPTG. Lane 3: molecular mass marker (200, 116.3, 97.4, 66.3, 55.4, 36.5, 31, 21.5, 14.4, 6 kDa kDa). The samples were incubated for 30 minutes at 37 °C before loading on the gel.

Fig. 4. The construction of the vector pMN501 for overexpression of MspA in *E. coli* BL21(DE3) is schematically depicted. The meaning of the abbreviations is as follows:

lacI: gene encoding the lactose repressor

nptI: gene encoding the neomycine phosphotransferase. This gene confers resistance against kanamycin.

Ori: origin of replication

RBS: ribosomal binding site

Fig. 5 shows schematically an apparatus for renaturation of monomeric MspA. A dispensable pipette tip from polyethylene of 5 cm length was shortened by 2 mm at its lower end. This tip was filled with a solution of 1.7% agarose in TAE buffer. The lead of a pencil (brand: Eberhard Faber, 3H) was shortened to 5 cm. A tube from polypropylene without a lid was filled with 60 µl of a solution containing 5 µg denatured MspA. The pipette tip and the lead were put into the solution and connected as cathode and anode, respectively.

Fig. 6 shows the renaturation of denatured MspA. The proteins were separated in a 10% denaturing polyacrylamide gel as described by Schägger (Schägger, H. and von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* 166, 368-79 (1987)). The gel was stained with silver. Lane M: molecular mass marker (116, 97, 66, 55, 36.5, 31, 21.5, 14.4kDa). Lane 1: 800 ng denatured MspA. Lane 2: 800 ng MspA after the renaturation reaction. The samples were incubated for 30 minutes at 37 °C before loading on the gel.

Fig. 7a through 7c show electron microscopic pictures of modifications of the channel-forming protein MspA from *M. smegmatis*. The preparation of the sample was performed as follows:

One milliliter of a solution of MspA (c(MspA)= 17.2 x 10^{-9} mol/L, 100 mM Na_2HPO_4/NaH_2PO_4, pH 6.5, 150 mM NaCl, 0.10 g/L SDS) was dispersed by ultrasonication in a water bath at 24.5°C. The distance between the liquid and HOPG (carbon) sur-

faces (1,0 mm^2) was 5.0 cm. The dispersed liquid droplets were allowed to contact the HOPG surface for 20 seconds.

Fig. 7a shows isolated channel-forming proteins. Fig. 7b shows a ribbon-like structure, which exhibits large pores with a diameter of 12 nm. Fig. 7c shows two types of channels in the ribbon-like structure: Channels with a small diameter of about 2.4 nm and channels with a larger diameter of about 9 to 10 nm.

Example 1: Extraction of MspA from *M. smegmatis* at different temperatures

Ten milligrams of *M. smegmatis* mc^2155 cells (wet weight) were washed with PBS (100 mM sodium phosphate, pH 7,0, 150 mM NaCl, 0.1 mM EDTA) and resuspended in 150 µl PG05 buffer (0.5% isotridecylpolyethylenglycolether, 100 mM Na_2HPO_4/NaH_2PO_4, 0.1 mM EDTA, 150 mM NaCl, pH 6.5). The resuspended cells were incubated for 30 minutes at 30, 40, 50, 60, 70, 80, 90 or 100°C. The samples were cooled on ice for 10 minutes and centrifuged for 10 minutes at 4°C. The volume of the supernatant was reduced from 120 µl to 10 µl by evaporation. The proteins were separated according to their sizes by gel electrophoresis as it is shown in the Figures 1a-c. The fraction of MspA compared to the total protein in the extract increases significantly at temperatures above 50°C. At those temperatures, only minor amounts of other proteins are extracted.

Example 2: Purification of MspA from *M. smegmatis*

Ten grams *M. smegmatis* cells were washed with PBS, resuspended in 35 mL POP05 and boiled under stirring for 30 minutes in a water bath. The cell suspension was cooled on ice for 10 minutes, and centrifuged at 4°C for 15 minutes at 27000 g. Forty-two milliliters of the supernatant were gently mixed with an

equal volume of ice-cold acetone. This mixture was kept on ice for one hour and centrifuged at 4°C for 15 minutes at 8000 g. The precipitated protein was dissolved in 10 mL 25mM *N*-(2-hydroxyethyl)piperazine-*N'*-2-ethane sulfonic acid (HEPES), pH 7.5, 10 mM NaCl, 0.5% OPOE (AOP05) and loaded on an anion exchange column "POROS 20HQ" with a volume of 1.7 mL (Perseptive Biosystems, Cambridge, USA). After washing the column with 14 mL AOP05, bound proteins were eluted with a gradient from 100% AOP05 to 100% BOP05 (25 mM HEPES, pH 7.5, 2 M NaCl, 0.5% OPOE) over 34 mL. Ninety fractions of 1 mL were collected and analysed by gel electrophoresis. Four fractions with the highest amount of MspA were pooled and the protein was precipitated with acetone as described above. The pellet was dissolved in 600 µL AOP05, incubated on ice and centrifuged at 4 °C for 5 minutes to remove insoluble material. The protein solution was loaded on a gel filtration column "Superdex G200" with a volume of 24 mL (Pharmacia, Freiburg, Germany). Proteins were eluted with 48 mL of AOP05 at a flow rate of 0.2 mL/min. Fifty fractions of 1 mL were collected and analyzed using denaturing polyacrylamide gels which were stained with silver. Fractions containing apparently pure MspA were pooled. The purification steps are shown in Fig. 2. The yield was 700 µg. 1 µg of this sample Probe did not show any contamination with other proteins in a silver-stained denaturing polyacrylamide gel (data not shown). Thus, MspA was purified to apparent homogeneity.

Example 3: Strategy for purification of the channel-forming protein MspA from *E. coli*

To further increase the yield of MspA, an overexpression of the *mspA* gene in *E. coli* is suggested. The *mspA* gene, which encodes the channel-forming protein MspA from *M. smegmatis*. The T7 expression system is chosen for overexpression of the *mspA* gene.

The *mspA* gene was amplified from the plasmid pPOR6 by PCR. All codons of the native mspA sequence, which occur rarely in highly expressed genes in *E. coli*, were exchanged. All mutations are listed in sequence 4 of the sequence protocol (see below). This gene was synthesized by assembling of oligonucleotides as described by Stemmer (Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M. and Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49-53 (1995)) and was called *synmspA*. The *synmspA* gene replaced the *mspA* gene in the vector pMN500, whose use did not lead to detectable amounts of MspA in *E. coli*, to give the vector pMN501 (Fig. 4). The vector pMN501 gave rise to a strong expression of the MspA monomer (20 kDa) in *E. coli* BL21(DE3) after induction with IPTG. This protein is called recombinant MspA (rMspA) and possesses the sequence 5 of the sequence protocol (see below).

Example 4: Procedure for purification of the channel-forming protein MspA from *E. coli*

One litre LB medium containing 30 µg/mL kanamycin is inoculated with mit *E. coli* BL21(DE3)/pMN501 and the culture was grown to an OD_{600} of 0.6 at 37 °C. Then, the cells are induced with 1 mM IPTG and are incubated at 37 °C for further six hours, until the culture reaches an OD_{600} of 2.2. The cells are harvested by centrifugation, resuspended in 40 mL A-Puffer (25 mM Hepes, pH 7,5, 10 mM NaCl) and lysed by boiling in water for 10 min. The cell lysate is kept on ice for 10 min and cell debris and insoluble proteins are precipitated by centrifugation at 10000 g for 10 min. The supernatant is fractionated using anion exchange chromatography (POROS HQ20, Perseptive Biosystems, Cambridge, USA) and a linear gradient from 10 mM to 2 M sodium chloride. Monomeric MspA elutes at 350 mM NaCl. The fractions containing MspA are pooled. Size-exclusion chromatography (Superdex G200, Pharmacia, Freiburg, Germany)

is used to purify MspA from proteins with a larger molecular weight. The yield is 10 mg MspA with a purity exceeding 95% (data not shown).

Example 5: Electrochemical assembly of the channel-forming protein MspA

Overexpression of MspA in *E. coli* easily allows to produce MspA with a good yield. However, a large fraction of the purified protein is inactive. Renaturation of MspA into its active form can be achieved by using the following protocol:
Renaturation can take place in an apparatus specially designed for this purpose (Fig. 5). The renaturation reaction is performed with 5 µg monomeric MspA in the aforementioned apparatus by applying a voltage of 50 V for 30 min. Then, the sign of the applied voltage is reversed for five seconds, to remove porin adsorbed at the surface of the lead. The protein is analysed in a denaturing denaturing polyacrylamide gel after the renaturation reaction (Fig. 6). This gel shows that a large fraction of the protein is assembled to oligomers. It is demonstrated by reconstitution in lipid bilayer experiments, that this assembled MspA has a high channel-forming activity. This experiment demonstrates that renaturation of monomeric MspA is possible using small DC voltages. This renaturation reaction is very easy to perform and is an important component of the purification of functional MspA from overexpressing *E. coli*.

List of sequences:

1. *mspA* gene, translated
2. *mspA* gene + promoter, translated
3. MspA protein with putative signal sequence
4. *synmspA* gene, translated
5. rMspA protein
6. *mspC* gene
7. MspC protein

8. *mspD* gene

9. *mspD* protein

SEQUENCE PROTOCOL

<110> Niederweis Dr., Michael

Bossmann Dr., Stefan

<120> Methods for the Production of a Channel-forming Protein

<130> 401172

<140>
<141>

<150> DE 199 43 520.0
<151> 1999-09-11

<150> DE 199 41 416.5
<151> 1999-08-31

<160> 9

<170> PatentIn Ver. 2.1

<210> 1
<211> 636
<212> DNA
<213> Mycobacterium smegmatis

<220>
<221> CDS
<222> (1)..(636)
<223> *mspA* gene

<400> 1

```
atg aag gca atc agt cgg gtg ctg atc gcg atg gtt gca gcc atc gcg   48
Met Lys Ala Ile Ser Arg Val Leu Ile Ala Met Val Ala Ala Ile Ala
  1               5                  10                  15

gcg ctt ttc acg agc aca ggc acc tct cac gca ggc ctg gac aac gag   96
Ala Leu Phe Thr Ser Thr Gly Thr Ser His Ala Gly Leu Asp Asn Glu
             20                  25                  30

ctg agc ctc gtt gat ggc cag gac cgc acc ctc acc gtg cag cag tgg   144
Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu Thr Val Gln Gln Trp
         35                  40                  45

gac acc ttc ctc aat ggt gtg ttc ccc ctg gac cgc aac cgt ctt acc   192
Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg Asn Arg Leu Thr
     50                  55                  60

cgt gag tgg ttc cac tcc ggt cgc gcc aag tac atc gtg gcc ggc ccc   240
Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr Ile Val Ala Gly Pro
 65                  70                  75                  80

ggt gcc gac gag ttc gag ggc acg ctg gaa ctc ggc tac cag atc ggc   288
Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly Tyr Gln Ile Gly
                 85                  90                  95

ttc ccg tgg tcg ctg ggt gtg ggc atc aac ttc agc tac acc acc ccg   336
Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser Tyr Thr Thr Pro
            100                 105                 110

aac atc ctg atc gac gac ggt gac atc acc gct ccg ccg ttc ggc ctg   384
```

```
Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala Pro Pro Phe Gly Leu
        115                 120                 125

aac tcg gtc atc acc ccg aac ctg ttc ccc ggt gtg tcg atc tcg gca   432
Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly Val Ser Ile Ser Ala
    130                 135                 140

gat ctg ggc aac ggc ccc ggc atc cag gaa gtc gca acg ttc tcg gtc   480
Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala Thr Phe Ser Val
145                 150                 155                 160

gac gtc tcc ggc gcc gag ggt ggc gtg gcc gtg tcg aac gcc cac ggc   528
Asp Val Ser Gly Ala Glu Gly Gly Val Ala Val Ser Asn Ala His Gly
                165                 170                 175

acc gtg acc ggt gcg gcc ggc ggt gtg ctg ctg cgt ccg ttc gcc cgc   576
Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg Pro Phe Ala Arg
            180                 185                 190

ctg atc gcc tcg acc ggt gac tcg gtc acc acc tac ggc gaa ccc tgg   624
Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr Gly Glu Pro Trp
        195                 200                 205

aac atg aac tga                                                   636
Asn Met Asn
    210
```

<210> 2
<211> 1423
<212> DNA
<213> Mycobacterium smegmatis

<220>
<221> -10_signal
<222> (323)..(328)
<223> putative promoter

<220>
<221> CDS
<222> (499)..(1134)
<223> *mspA* gene

<220>
<221> RBS
<222> (492)..(496)
<223> putative ribosome binding site

<400> 2
gttaacggag tcgggccgtc gatacggcgg cgaagatcat ccggcagatt ggcgcctggt 60

taaacccgcg taaacactgg taccgccggt ccgcgccgga aaaggttttg cctcacggtg 120

aatatgtgac ctgaattgca cttcacgggt aaaagcggag gtaaccgacg gttgccgcag 180

caccctcaca gcttgggcca aggtgacgtg cagcgcacgc ctgccggtgc cggatggcgg 240

tcaccgcaaa gtgtcaggca ctgccgaaag gtcagtcagc aaacttcact gcggctgtgg 300

tgcgaagtgc ggttgtggga cgtatccgtt gctgccgcgc gccctggcgt ttatgtttct 360

gctgccaact gtgagcgagg cattagagac agatgtgatc ctcttagatc tccgaagtct 420

```
ctgaacaggt gttgagccgg ttgcagacaa caaaacaggt gggcctgagg ggccgccggc 480

gatacagtta gggagaac atg aag gca atc agt cgg gtg ctg atc gcg atg   531
                    Met Lys Ala Ile Ser Arg Val Leu Ile Ala Met
                      1               5                  10

gtt gca gcc atc gcg gcg ctt ttc acg agc aca ggc acc tct cac gca   579
Val Ala Ala Ile Ala Ala Leu Phe Thr Ser Thr Gly Thr Ser His Ala
             15                  20                  25

ggc ctg gac aac gag ctg agc ctc gtt gat ggc cag gac cgc acc ctc   627
Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
         30                  35                  40

acc gtg cag cag tgg gac acc ttc ctc aat ggt gtg ttc ccc ctg gac   675
Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
     45                  50                  55

cgc aac cgt ctt acc cgt gag tgg ttc cac tcc ggt cgc gcc aag tac   723
Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr
 60                  65                  70                  75

atc gtg gcc ggc ccc ggt gcc gac gag ttc gag ggc acg ctg gaa ctc   771
Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu
                 80                  85                  90

ggc tac cag atc ggc ttc ccg tgg tcg ctg ggt gtg ggc atc aac ttc   819
Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe
             95                  100                 105

agc tac acc acc ccg aac atc ctg atc gac gac ggt gac atc acc gct   867
Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala
         110                 115                 120

ccg ccg ttc ggc ctg aac tcg gtc atc acc ccg aac ctg ttc ccc ggt   915
Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly
     125                 130                 135

gtg tcg atc tcg gca gat ctg ggc aac ggc ccc ggc atc cag gaa gtc   963
Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val
140                 145                 150                 155

gca acg ttc tcg gtc gac gtc tcc ggc gcc gag ggt ggc gtg gcc gtg   1011
Ala Thr Phe Ser Val Asp Val Ser Gly Ala Glu Gly Gly Val Ala Val
                160                 165                 170

tcg aac gcc cac ggc acc gtg acc ggt gcg gcc ggc ggt gtg ctg ctg   1059
Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu
            175                 180                 185

cgt ccg ttc gcc cgc ctg atc gcc tcg acc ggt gac tcg gtc acc acc   1107
Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr
        190                 195                 200

tac ggc gaa ccc tgg aac atg aac tga ttcctggacc gccgttcggt         1154
Tyr Gly Glu Pro Trp Asn Met Asn
    205                 210

cgctgagacc gcttgagatc ggcgcgtccc gctcccggtg tcgtcagctc atcgttgaca 1214

cgtgaactga cactcttcct agccggagcg kacgcgccga tcttgtgttc tgagcagttc 1274

tcagtccgtc cgccgcaaca ccagcgctga cggcgtacgc agcctgccca ccaccgcgcg 1334
```

ccagggacgc cccagcctgg gcaccacctc agcggtcggc acgatgcgcg gatcggtcac 1394

ctcgaacgtc tcaccgttca tcaccgcgc 1423

<210> 3
<211> 211
<212> PRT
<213> Mycobacterium smegmatis

<220>
<221> signal
<222> (1)..(27)
<223> putative signal sequence of the MspA protein

<220>
<221> PEPTIDE
<222> (28)..(211)
<223> mature MspA protein

<400> 3

```
Met Lys Ala Ile Ser Arg Val Leu Ile Ala Met Val Ala Ala Ile Ala
 1               5                  10                  15

Ala Leu Phe Thr Ser Thr Gly Thr Ser His Ala Gly Leu Asp Asn Glu
            20                  25                  30

Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu Thr Val Gln Gln Trp
        35                  40                  45

Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg Asn Arg Leu Thr
    50                  55                  60

Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr Ile Val Ala Gly Pro
65                  70                  75                  80

Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly Tyr Gln Ile Gly
                85                  90                  95

Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser Tyr Thr Thr Pro
            100                 105                 110

Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala Pro Pro Phe Gly Leu
        115                 120                 125

Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly Val Ser Ile Ser Ala
    130                 135                 140

Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala Thr Phe Ser Val
145                 150                 155                 160

Asp Val Ser Gly Ala Glu Gly Gly Val Ala Val Ser Asn Ala His Gly
                165                 170                 175

Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg Pro Phe Ala Arg
            180                 185                 190

Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr Gly Glu Pro Trp
        195                 200                 205

Asn Met Asn
    210
```

<210> 4
<211> 558
<212> DNA
<213> synthetic sequence

<220>
<221> CDS
<222> (1)..(558)
<223> *synmspA* gene

<400> 4

```
atg ggc ctg gac aac gaa ctg tcc ctg gtt gac ggc cag gac cgt acc   48
Met Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr
  1               5                  10                  15

ctg acc gtt cag cag tgg gac acc ttc ctg aac ggt gtt ttc ccg ctg   96
Leu Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu
             20                  25                  30

gac cgt aac cgt ctg acc cgt gaa tgg ttc cac tcc ggt cgt gcg aaa   144
Asp Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys
         35                  40                  45

tac atc gtt gcg ggt ccg ggt gcg gac gag ttc gaa ggt acc ctg gaa   192
Tyr Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu
     50                  55                  60

ctg ggt tac cag atc ggc ttc ccg tgg tcc ctg ggt gtt ggt atc aac   240
Leu Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn
 65                  70                  75                  80

ttc tct tac acc acc ccg aac atc ctg atc gac gac ggt gac atc acc   288
Phe Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr
                 85                  90                  95

gct ccg ccg ttc ggt ctg aac tct gtt atc acc ccg aac ctg ttc ccg   336
Ala Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro
            100                 105                 110

ggt gtt tct atc tct gct gat ctg ggc aac ggt ccg ggt atc cag gaa   384
Gly Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu
        115                 120                 125

gtt gct acc ttc tct gta gac gtc tct ggt gct gaa ggt ggt gtt gct   432
Val Ala Thr Phe Ser Val Asp Val Ser Gly Ala Glu Gly Gly Val Ala
    130                 135                 140

gtt tct aac gct cac ggc acc gtt acc ggt gcg gct ggc ggt gtt ctg   480
Val Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu
145                 150                 155                 160

ctg cgt ccg ttc gct cgt ctg atc gct tct acc ggt gac tct gtt acc   528
Leu Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr
                165                 170                 175

acc tac ggt gaa ccg tgg aac atg aac tga                          558
Thr Tyr Gly Glu Pro Trp Asn Met Asn
            180                 185
```

<210> 5
<211> 185
<212> PRT
<213> synthetic sequence

<220>
<221> peptide
<222> (1)..(184)
<223> rMspA

<220>
<223> description of the synthetic sequence:synthetic

<400> 5
Met Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr
1 5 10 15

Leu Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu
20 25 30

Asp Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys
35 40 45

Tyr Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu
50 55 60

Leu Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn
65 70 75 80

Phe Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr
85 90 95

Ala Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro
100 105 110

Gly Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu
115 120 125

Val Ala Thr Phe Ser Val Asp Val Ser Gly Ala Glu Gly Gly Val Ala
130 135 140

Val Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu
145 150 155 160

Leu Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr
165 170 175

Thr Tyr Gly Glu Pro Trp Asn Met Asn
180 185

<210> 6
<211> 648
<212> DNA
<213> Mycobacterium smegmatis

<400> 6
atgaaggcaa tcagtcgggt gctgatcgcg atgatttccg cgttggctgc ggccgtcgcg 60
gggttgttcg tgagcgcggg cacctctcac gcgggtctcg acaatgagct cagccttgtc 120
gatggtcagg accgcaccct caccgtgcag cagtgggata cgttcctcaa tggtgtgttc 180
cccctggacc gcaaccgtct gacccgtgag tggttccact ccggtcgcgc gaagtacatc 240
gtggccggcc ccggtgccga tgagttcgag ggcacgctgg aactcggcta ccagatcggc 300
ttcccgtggt cgctgggtgt gggcatcaac ttcagctaca ccaccccgaa catcctgatc 360

gacgacggtg acatcaccgg tccgcccttc ggcctcgagt cggtcatcac cccgaacctg 420
ttccccggtg tgtcgatctc ggccgacctg ggcaacggcc ccggcatcca ggaagtcgcg 480
acgttctcgg tcgacgtctc gggtcccgca ggcggagtag cggtctccaa cgcgcacggc 540
accgtcaccg gtgcggccgg cggtgtgctg ctgcgtccgt tcgcccgcct gatcgcctcg 600
accggtgact cggtcaccac ctacggcgaa ccctggaaca tgaactga 648

<210> 7
<211> 184
<212> PRT
<213> Mycobacterium smegmatis

<400> 7
Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
1 5 10 15

Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
20 25 30

Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr
35 40 45

Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu
50 55 60

Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe
65 70 75 80

Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly
85 90 95

Pro Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly
100 105 110

Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val
115 120 125

Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val
130 135 140

Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu
145 150 155 160

Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr
165 170 175

Tyr Gly Glu Pro Trp Asn Met Asn
180

<210> 8
<211> 624
<212> DNA
<213> Mycobacterium smegmatis

<400> 8
gtgcgctacc tcgtcatgat gttcgctcta ctcgtgtccg tgacgcttgt gagcccccgc 60
cctgccaacg cggtggacaa tcagctcagc gtggtcgacg gccaaggtcg cacgctgacc 120
gtgcagcaag ccgagacatt cctcaacggc gtgttccctc tcgaccggaa ccgactgacc 180
cgtgagtggt ttcactccgg ccgcgccacc taccatgtgg ccggcccagg tgccgacgaa 240
ttcgagggca cgctcgaact cgggtatcag gtcggcttcc cgtggtcatt gggcgtcggc 300
atcaacttct cgtacacgac cccgaacatc ctcatcgacg gaggcgacat cacccagccg 360
ccgttcggcc tggacaccat catcaccccc aacctcttcc ccggcgtgtc catcagtgcc 420

gacctcggca acggtcccgg tatccaggag gtcgccacct tctcggtgga cgtgaagggc 480
gcgaaaggag cggtcgccgt atccaatgcg catggcaccg tgaccggcgc ggccggcggc 540
gtgctcctgc gtccgttcgc ccggttgatc gcctcgacgg gcgacagcgt caccacctac 600
ggcgagccct ggaacatgaa ctag 624

<210> 9
<211> 183
<212> PRT
<213> Mycobacterium smegmatis

<400> 9

Val	Asp	Asn	Gln	Leu	Ser	Val	Val	Asp	Gly	Gln	Gly	Arg	Thr	Leu	Thr
1				5				10					15		
Val	Gln	Gln	Ala	Glu	Thr	Phe	Leu	Asn	Gly	Val	Phe	Pro	Leu	Asp	Arg
		20					25					30			
Asn	Arg	Leu	Thr	Arg	Glu	Trp	Phe	His	Ser	Gly	Arg	Ala	Thr	Tyr	His
	35					40					45				
Val	Ala	Gly	Pro	Gly	Ala	Asp	Glu	Phe	Glu	Gly	Thr	Leu	Glu	Leu	Gly
50					55					60					
Tyr	Gln	Val	Gly	Phe	Pro	Trp	Ser	Leu	Gly	Val	Gly	Ile	Asn	Phe	Ser
65				70					75					80	
Tyr	Thr	Thr	Pro	Asn	Ile	Leu	Ile	Asp	Gly	Gly	Asp	Ile	Thr	Gln	Pro
			85					90					95		
Pro	Phe	Gly	Leu	Asp	Thr	Ile	Ile	Thr	Pro	Asn	Leu	Phe	Pro	Gly	Val
		100					105					110			
Ser	Ile	Ser	Ala	Asp	Leu	Gly	Asn	Gly	Pro	Gly	Ile	Gln	Glu	Val	Ala
	115					120					125				
Thr	Phe	Ser	Val	Asp	Val	Lys	Gly	Ala	Lys	Gly	Ala	Val	Ala	Val	Ser
130					135					140					
Asn	Ala	His	Gly	Thr	Val	Thr	Gly	Ala	Ala	Gly	Gly	Val	Leu	Leu	Arg
145				150					155					160	
Pro	Phe	Ala	Arg	Leu	Ile	Ala	Ser	Thr	Gly	Asp	Ser	Val	Thr	Thr	Tyr
			165					170					175		
Gly	Glu	Pro	Trp	Asn	Met	Asn									
		180													